

Joint Inventors

Docket No. INTEL/18500
P18500

"EXPRESS MAIL" mailing label No.
EV 309992111 US
Date of Deposit: **February 20, 2004**

I hereby certify that this paper (or fee) is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 CFR § 1.10 on the date indicated above and is addressed to:
Commissioner for Patents, P.O. Box 1450,
Alexandria, VA 22313-1450


Charissa D. Wheeler

APPLICATION FOR UNITED STATES LETTERS PATENT

SPECIFICATION

TO ALL WHOM IT MAY CONCERN:

Be it known that We, **Gururaj Nagendra**, a citizen of India, residing at 1441 NE Carlabay Way, #94, Hillsboro, Oregon 97124; and **Stewart Taylor**, a citizen of United States, residing at 1202 Sharon Park Dr., #69, Menlo Park, California 94025 have invented a new and useful **METHODS AND APPARATUS TO OPTIMIZE MANAGED APPLICATION PROGRAM INTERFACES**, of which the following is a specification.

**METHODS AND APPARATUS TO OPTIMIZE MANAGED APPLICATION
PROGRAM INTERFACES**

TECHNICAL FIELD

[0001] The present disclosure relates generally to managed runtime environments, and more particularly, to methods and apparatus to optimize managed application program interfaces (APIs).

BACKGROUND

[0002] Managed code is code executing under the control of a managed runtime environment (MRTE) (e.g., any code written in C# (“C-sharp”) from Microsoft® or Visual Basic .NET), whereas unmanaged code is code executing outside of the MRTE (e.g., COM components and WIN32 API functions). Typically, managed code may be used to support components and applications during runtime, and unmanaged code may be used to support low-level interaction with the platform (i.e., the processor). As applications migrate toward operability on MRTEs such as Java® Virtual Machine (JVM) and Common Language Runtime (CLR) provided by Microsoft® .NET, virtual machines are abstracting the applications away from processors (i.e., managed runtime applications are becoming more dependent on the virtual machines and less dependent on the processors).

[0003] Currently, unmanaged software library functions such as Intel® Integrated Performance Primitives (IPP) are generally optimized for execution in unmanaged environments on processors implemented using one or more of the Intel® Pentium® technology and/or the Intel® Itanium® technology. The unmanaged software library functions may be further optimized to operate on a specific processor architecture by writing custom hand optimization code with processor-specific instructions such as a

Streaming Single Instruction/Multiple Data (SIMD) Extension (SSE) instruction, an SSE2 instruction, and/or a MultiMedia Extension (MMX) instruction offered by Intel[®] processors. For example, a String Compare function may be implemented in unmanaged code and optimized by custom hand optimization coding using the SSE2 instruction. In contrast to unmanaged code, managed code may not be optimized for particular processor architectures in the same way as unmanaged code because no mechanism exists to custom hand optimize managed code. For example, typically, managed APIs are solely dependent on a just-in-time (JIT) compiler for optimization. As a result, managed runtime applications are unable to take advantage of processor-specific optimizing instructions for execution on an underlying processor to enable and optimize features such as audio processing, video processing, image processing, speech recognition, cryptography, etc.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] FIG. 1 is a block diagram representation of an example architectural hierarchy of a managed runtime environment (MRTE) system configured in an existing system.

[0005] FIG. 2 is a block diagram representation of an example architectural hierarchy of an example MRTE system including a processor instruction proxy stubs (PIPS) system configured in accordance with an embodiment of the teachings of the invention as disclosed herein.

[0006] FIG. 3 is a block diagram representation of an example processor instruction proxy stubs (PIPS) system.

[0007] FIG. 4 is a high level language representation of example unmanaged code that may be optimized by an example PIPS system as in FIG. 3.

[0008] FIG. 5 is a code representation of example native assembly code corresponding to the high level language of FIG. 4 and including a PIPS that optimizes the native assembly code.

[0009] FIGS. 6 and 7 are flow diagram representations of example machine accessible instructions that may be executed to implement an example PIPS system as in FIG. 3.

[0010] FIG. 8 is a block diagram representation of an example processor system that may be used to implement an example PIPS system as in FIG. 3.

DETAILED DESCRIPTION

[0011] Referring to FIG. 1, an architectural hierarchy of a managed runtime environment (MRTE) system 100 typically includes a managed runtime application 110, one or more managed application program interfaces (APIs) 120, a virtual machine (VM) 130, a compiler 140, processor-specific instructions 150, and a processor 160. As used herein the term “application” refers to one or more methods, programs, functions, routines, or subroutines for manipulating data.

[0012] Typically, the managed runtime application 110 is written by programmers to provide various services in an MRTE. The source code of the managed runtime application 110 may be written in, for example, C#, Visual Basic .NET, and/or any other suitable object-oriented programming languages.

[0013] The managed APIs 120 such as Microsoft® .NET Framework Class Libraries or Java Class Libraries convert (i.e., compile) the source code of the managed runtime application 110 into Microsoft Intermediate Language (MSIL) code or Java byte code, respectively. The managed APIs 120 serve as an interface between the managed runtime application 110 and the VM 130.

[0014] The VM 130 operates an abstract processor to manage the managed runtime application 110 by providing services such as garbage collection, memory management, and code and role-based security to the managed APIs 120. For example, the VM 130, which is processor agnostic, may be a Microsoft Common Language Runtime or a Java Virtual Machine. The managed APIs 120 and the VM 130 operate independent of any specific platform so that the MISL code or the Java byte code is not targeted to any specific processor. Accordingly, the compiler 140 such as a just-in-time (JIT) compiler converts (i.e., re-compiles) the MISL code or the Java byte code from the managed APIs 120 into native assembly code that may be executed by the processor 160.

[0015] The processor 160 may be implemented using one or more of the Intel® Pentium® technology, the Intel® Itanium® technology, and/or Intel® Personal Internet Client Architecture (PCA) technology. The processor 160 may be capable of executing processor-specific instructions 150 such as SSE instructions, SSE2 instructions, MMX instructions and/or other suitable instructions to provide software library functions such as cryptography, multimedia, audio codecs, video codecs, image coding, image processing, signal processing, string processing, speech compression, computer vision, etc. to the MRTE system 100.

[0016] As mentioned above, however, unmanaged software library functions (i.e., processor-specific instructions 150) may be optimized for the processor 160 whereas managed code (i.e., the managed APIs 120) may not be optimized for certain processor architectures in the same way because previously no mechanism exists to custom-hand optimize managed code functions. That is, the managed APIs 120 corresponding to the managed runtime application 110 were solely dependent on the JIT compiler 140 for optimization and the JIT compiler 140 was incapable of processor-specific optimization. Thus, in prior systems, the underlying processor 160 was not able to take advantage of the

services provided by the VM 130, while the managed runtime application 110 was not able to take advantage of the features provided by the underlying processor 160 because the VM 130 did not support certain processor-specific instructions 150 of the underlying processor 160.

[0017] In the example of FIG. 2, an illustrated architectural hierarchy of an MRTE including a processor instruction proxy stub (PIPS) system 200 includes a managed runtime application 210, one or more APIs 220, one or more optimized managed APIs 225, a VM 230, a PIPS generator 235, a compiler 240, processor-specific instructions 250, and a processor 260. As used herein “stub” refers to a portion of dynamically-generated code provided to perform various tasks during execution of a program.

[0018] In general, the PIPS generator 235 generates a portion of code or set of instructions referred to as a PIPS (e.g., PIPS 510 of FIG. 5) to optimize execution of the managed runtime application 210 on the underlying processor 260. When the managed runtime application 210 is installed, for example, the PIPS generator 235 generates a PIPS based on the processor-specific instructions 250. Further, the PIPS generator 235 inserts the PIPS into certain managed APIs 220 to create the optimized managed APIs 225 used by the managed runtime application 210. During execution of the managed runtime application 210 as described in detail below, the optimized managed APIs 225 optimize performance of the underlying processor 260 without having to rewrite unmanaged code (i.e., the processor-specific instructions 250) to managed code (i.e., the managed runtime application 210). The optimized managed APIs 225 may be stored in memory (e.g., memory 1030 of FIG. 8) and recalled during execution of the managed runtime application 210 in an MRTE. As a result, the features of the underlying processor 260 may be enabled to optimize performance of the managed runtime application 210 on the underlying processor 260.

[0019] While the PIPS generator 235 shown in FIG. 2 is depicted as a separate block within the PIPS system 200, the functions performed by the PIPS generator 235 may be integrated within the VM 230 and/or the JIT compiler 240.

[0020] Referring to FIG. 3, an example PIPS system 300 includes a managed runtime application 310, one or more optimized managed APIs 325, a VM 330, a JIT compiler 340, native assembly code 350, and a processor 360 to execute the managed runtime application 310 in an MRTE. The VM 330 may execute processor instructions compatible with different processors to execute the managed runtime application 310. Typically, however, the VM 330 may not execute certain processor-specific instructions of the underlying processor 360 to enable features that would otherwise be unavailable without the optimized managed APIs 325. In contrast, during execution of the managed runtime application 310 by the PIPS system 300, for example, the JIT compiler 340 compiles the optimized managed APIs 325 to generate the native assembly code 350 (e.g., the native assembly code 500 of FIG. 5). In particular, the JIT compiler 340 simply compiles and executes the native assembly code 350 without having to optimize the native assembly code 350 any further because the PIPS generator 235 inserted the PIPS to generate the optimized managed APIs 325 during installation of the managed runtime application 310. In other words, the PIPS previously optimized the managed APIs of the managed runtime application 310 (i.e., the optimized managed APIs 325) for the execution of the managed runtime application 310 on the underlying processor 360. Accordingly, the optimized managed APIs 325 optimize performance of the underlying processor 360 without the JIT compiler 340 rewriting unmanaged code (e.g., processor-specific instructions 250 of FIG. 2) to managed code (i.e., the managed runtime application 310). As a result, the native assembly code 350 is customized to optimize performance of the managed runtime application 310 on the underlying processor 360.

[0021] In the example of FIG. 4, a String Compare function 400 is implemented in unmanaged high-level code. Typically, the String Compare function 400 is optimized as a C language routine by custom-hand optimized coding using processor-specific instructions such as SSE2 instructions for a processor implemented using one or more of the Intel® processing technology mentioned above. However, no mechanism exists to custom-hand optimize managed code such as C# or Java Compare function code for a particular processing architecture.

[0022] As described in conjunction with FIGS. 2 and 3, an example portion of native assembly code 500 including a PIPS 510 to optimize the performance of the String Compare function 400 on the underlying processor 360 is shown in FIG. 5. In particular, the native assembly code 500 includes a PIPS 510 generated by the PIPS generator 235. For example, the PIPS generator 235 may use native marshaling language (ML) code provided by Microsoft® .NET to generate the PIPS 510 during installation of the String Compare function 400. Based on the PIPS 510, the PIPS generator 235 creates the optimized managed APIs 325 corresponding to the managed runtime application 310. The JIT compiler 340 compiles the native assembly code 500 corresponding to the String Compare function, which includes the PIPS 510, as shown in FIG. 5 for the underlying processor 360 to execute. When the String Compare function 400 is initiated during runtime, the VM 330 retrieves the optimized managed APIs 325 for the JIT compiler 340 to generate the native assembly code 500. The JIT compiler 340 compiles and executes optimized managed APIs 325 without having to optimize the optimized managed APIs 325 any further because the PIPS generator 235 previously inserted the PIPS 510 into the optimized managed APIs 325. As a result, the managed runtime application 310 may benefit from both the services provided by the VM 330 (e.g., garbage collection, memory management, and/or code and role-based security) and the features of the underlying

processor 360 because the processor-specific instructions 250 (i.e., unmanaged code) of the underlying processor 360 are abstracted up to the VM layer via the PIPS 510. In other words, the optimized managed APIs 325 may enable processor-specific instructions to enable features of the underlying processor 360 to operate the managed runtime applications 310.

[0023] Flow diagrams 600 and 700 representing machine accessible instructions that may be executed by a processor to optimize managed APIs are illustrated in FIGS. 6 and 7, respectively. Persons of ordinary skill in the art will appreciate that the instructions may be implemented in any of many different ways utilizing any of many different programming codes stored on any of many computer-accessible mediums such as a volatile or nonvolatile memory or other mass storage device (e.g., a floppy disk, a CD, and a DVD). For example, the machine accessible instructions may be embodied in a machine-accessible medium such as an erasable programmable read only memory (EPROM), a read only memory (ROM), a random access memory (RAM), a magnetic media, an optical media, and/or any other suitable type of medium. Alternatively, the machine accessible instructions may be embodied in a programmable gate array and/or an application specific integrated circuit (ASIC). Further, although a particular order of actions is illustrated in FIGS. 6 and 7, persons of ordinary skill in the art will appreciate that these actions can be performed in other temporal sequences. Again, the flow diagrams 600 and 700 are merely provided and described in conjunction with FIGS. 2 and 5 as an example of one way to optimize managed APIs.

[0024] In the example of FIG. 6, the flow diagram 600 begins with the PIPS generator 235 generating the PIPS 510 associated with processor-specific instructions 250 of the underlying processor 260 (block 610). For example, the PIPS generator 235 may generate the PIPS 510 based on a processor identifier corresponding to the underlying

processor 260 during installation of the managed runtime application 210. As noted above, the processor-specific instructions 250 enable features of the underlying processor 260 such as audio processing, video processing, image processing, speech recognition, cryptography, etc. to optimize performance of the managed runtime application 210 on the underlying processor 260 when such features may be otherwise unavailable. Based on the PIPS 510, the PIPS generator 235 generates the optimized managed APIs 225 (block 620). In particular, the PIPS generator 235 inserts the PIPS 510 into certain managed APIs 220 corresponding to the managed runtime application 210. The PIPS generator 235 stores the optimized managed APIs 225 so that the optimized managed APIs 235 may be available for the JIT compiler 240 during execution of the managed runtime application 210 on the underlying processor 260.

[0025] In the example of FIG. 7, a flow diagram 700 begins with the JIT compiler 240 compiling and executing the optimized managed APIs 225 corresponding to the managed runtime application 210 (block 710). As noted above, the JIT compiler 240 may compile the optimized managed APIs 225 without further optimizing the optimized managed APIs 225 because the PIPS generator 235 previously inserted the PIPS 510 associated with the processor-specific instructions 250 into the optimized managed APIs 225. That is, the PIPS 510 custom-hand optimizes the managed runtime application 210 to operate on the underlying processor 260 via the optimized managed APIs 225. The JIT compiler 240 enables features of the underlying processor 260 corresponding to the processor-specific instructions 250 (block 640). In addition to services such as garbage collection, memory management, and code and role-based security provided by the VM 230, the managed runtime application 210 may take advantage of the software library functions provided by the optimized managed APIs 225 such as cryptography, multimedia, audio codecs, video codecs, image coding, image processing, signal

processing, string processing, speech compression, computer vision, etc. to the managed runtime application 210 during execution on the underlying processor 260. As a result, the managed optimized APIs 225 permit the managed runtime application 210 to execute processor-specific instructions 250 to enable features of the underlying processor 260 that otherwise would be unavailable or inefficient on another processor. Further, the managed optimized APIs 225 custom-hand optimizes performance of the managed runtime application 210 on the underlying processor 260 via the native assembly code 500.

[0026] The methods and apparatus disclosed herein are well suited for source code to implementations of the European Computer Management Association (ECMA) Common Language Infrastructure (CLI) (second edition, December 2002) and the ECMA C# language specification (second edition, December 2002). However, persons of ordinary skill in the art will appreciate that the teachings of the disclosure may be applied to source code in other runtime environments.

[0027] FIG. 8 is a block diagram of an example processor system 1000 adapted to implement the methods and apparatus disclosed herein. The processor system 1000 may be a desktop computer, a laptop computer, a notebook computer, a personal digital assistant (PDA), a server, an Internet appliance or any other type of computing device.

[0028] The processor system 1000 illustrated in FIG. 8 includes a chipset 1010, which includes a memory controller 1012 and an input/output (I/O) controller 1014. As is well known, a chipset typically provides memory and I/O management functions, as well as a plurality of general purpose and/or special purpose registers, timers, etc. that are accessible or used by a processor 1020. The processor 1020 is implemented using one or more processors. For example, the processor 1020 may be implemented using one or more of the Intel® Pentium® technology, the Intel® Itanium® technology, Intel® Centrino™ technology, and/or the Intel® XScale® technology. In the alternative, other

processing technology may be used to implement the processor 1020. The processor 1020 includes a cache 1022, which may be implemented using a first-level unified cache (L1), a second-level unified cache (L2), a third-level unified cache (L3), and/or any other suitable structures to store data as persons of ordinary skill in the art will readily recognize.

[0029] As is conventional, the memory controller 1012 performs functions that enable the processor 1020 to access and communicate with a main memory 1030 including a volatile memory 1032 and a non-volatile memory 1034 via a bus 1040. The volatile memory 1032 may be implemented by Synchronous Dynamic Random Access Memory (SDRAM), Dynamic Random Access Memory (DRAM), RAMBUS Dynamic Random Access Memory (RDRAM), and/or any other type of random access memory device. The non-volatile memory 1034 may be implemented using flash memory, Read Only Memory (ROM), Electrically Erasable Programmable Read Only Memory (EEPROM), and/or any other desired type of memory device.

[0030] The processor system 1000 also includes an interface circuit 1050 that is coupled to the bus 1040. The interface circuit 1050 may be implemented using any type of well known interface standard such as an Ethernet interface, a universal serial bus (USB), a third generation input/output interface (3GIO) interface, and/or any other suitable type of interface.

[0031] One or more input devices 1060 are connected to the interface circuit 1050. The input device(s) 1060 permit a user to enter data and commands into the processor 1020. For example, the input device(s) 1060 may be implemented by a keyboard, a mouse, a touch-sensitive display, a track pad, a track ball, an isopoint, and/or a voice recognition system.

[0032] One or more output devices 1070 are also connected to the interface circuit 1050. For example, the output device(s) 1070 may be implemented by display devices (e.g., a light emitting display (LED), a liquid crystal display (LCD), a cathode ray tube (CRT) display, a printer and/or speakers). The interface circuit 1050, thus, typically includes, among other things, a graphics driver card.

[0033] The processor system 1000 also includes one or more mass storage devices 1080 to store software and data. Examples of such mass storage device(s) 1080 include floppy disks and drives, hard disk drives, compact disks and drives, and digital versatile disks (DVD) and drives.

[0034] The interface circuit 1050 also includes a communication device such as a modem or a network interface card to facilitate exchange of data with external computers via a network. The communication link between the processor system 1000 and the network may be any type of network connection such as an Ethernet connection, a digital subscriber line (DSL), a telephone line, a cellular telephone system, a coaxial cable, etc.

[0035] Access to the input device(s) 1060, the output device(s) 1070, the mass storage device(s) 1080 and/or the network is typically controlled by the I/O controller 1014 in a conventional manner. In particular, the I/O controller 1014 performs functions that enable the processor 1020 to communicate with the input device(s) 1060, the output device(s) 1070, the mass storage device(s) 1080 and/or the network via the bus 1040 and the interface circuit 1050.

[0036] While the components shown in FIG. 8 are depicted as separate blocks within the processor system 1000, the functions performed by some of these blocks may be integrated within a single semiconductor circuit or may be implemented using two or more separate integrated circuits. For example, although the memory controller 1012 and the I/O controller 1014 are depicted as separate blocks within the chipset 1010, persons of

ordinary skill in the art will readily appreciate that the memory controller 1012 and the I/O controller 1014 may be integrated within a single semiconductor circuit.

[0037] Although certain example methods, apparatus, and articles of manufacture have been described herein, the scope of coverage of this patent is not limited thereto. On the contrary, this patent covers all methods, apparatus, and articles of manufacture fairly falling within the scope of the appended claims either literally or under the doctrine of equivalents. For example, although the above discloses example systems including, among other components, software or firmware executed on hardware, it should be noted that such systems are merely illustrative and should not be considered as limiting. In particular, it is contemplated that any or all of the disclosed hardware, software, and/or firmware components could be embodied exclusively in hardware, exclusively in software, exclusively in firmware or in some combination of hardware, software, and/or firmware.